

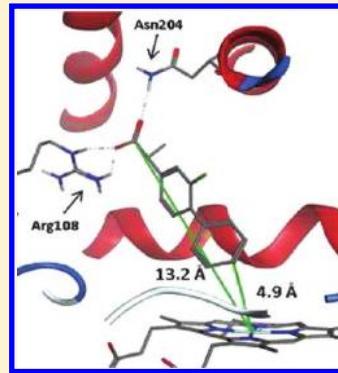
2D SMARTCyp Reactivity-Based Site of Metabolism Prediction for Major Drug-Metabolizing Cytochrome P450 Enzymes

Ruifeng Liu,*† Jin Liu,† Greg Tawa,† and Anders Wallqvist†

†DoD Biotechnology High Performance Computing Software Applications Institute, Telemedicine and Advanced Technology Research Center, U.S. Army Medical Research and Material Command, Fort Detrick, Maryland 21702, United States

 Supporting Information

ABSTRACT: Cytochrome P450 (CYP) 3A4, 2D6, 2C9, 2C19, and 1A2 are the most important drug-metabolizing enzymes in the human liver. Knowledge of which parts of a drug molecule are subject to metabolic reactions catalyzed by these enzymes is crucial for rational drug design to mitigate ADME/toxicity issues. SMARTCyp, a recently developed 2D ligand structure-based method, is able to predict site-specific metabolic reactivity of CYP3A4 and CYP2D6 substrates with an accuracy that rivals the best and more computationally demanding 3D structure-based methods. In this article, the SMARTCyp approach was extended to predict the metabolic hotspots for CYP2C9, CYP2C19, and CYP1A2 substrates. This was accomplished by taking into account the impact of a key substrate-receptor recognition feature of each enzyme as a correction term to the SMARTCyp reactivity. The corrected reactivity was then used to rank order the likely sites of CYP-mediated metabolic reactions. For 60 CYP1A2 substrates, the observed major sites of CYP1A2 catalyzed metabolic reactions were among the top-ranked 1, 2, and 3 positions in 67%, 80%, and 83% of the cases, respectively. The results were similar to those obtained by MetaSite and the reactivity + docking approach. For 70 CYP2C9 substrates, the observed sites of CYP2C9 metabolism were among the top-ranked 1, 2, and 3 positions in 66%, 86%, and 87% of the cases, respectively. These results were better than the corresponding results of StarDrop version 5.0, which were 61%, 73%, and 77%, respectively. For 36 compounds metabolized by CYP2C19, the observed sites of metabolism were found to be among the top-ranked 1, 2, and 3 sites in 78%, 89%, and 94% of the cases, respectively. The computational procedure was implemented as an extension to the program SMARTCyp 2.0. With the extension, the program can now predict the site of metabolism for all five major drug-metabolizing enzymes with an accuracy similar to or better than that achieved by the best 3D structure-based methods. Both the Java source code and the binary executable of the program are freely available to interested users.



1. INTRODUCTION

The cytochrome P450 superfamily of enzymes (abbreviated as CYP) comprises a large and diverse group of proteins with the heme cofactor.¹ In humans, they transform lipophilic drugs to more polar compounds that can be excreted by the kidneys and, therefore, play important roles in defining a drug's pharmacokinetic profile.² They also contribute to drug–drug interactions and metabolism-dependent toxicity issues.³ Among the CYP enzymes, CYP3A4, 2D6, 2C9, 2C19, and 1A2 are the most important members in human liver, the principal organ for phase 1 metabolism and clearance of drugs and other chemicals.⁴ Together, the five enzymes metabolize approximately 90% of marketed drugs.^{5–7} The ability to predict substrate specific sites of metabolism (SOM) of these enzymes is essential for rational drug design to mitigate ADME and toxicity issues of a drug candidate.

The mechanism of CYP-mediated metabolism is complex and consists of multiple-steps.⁸ However, there is experimental evidence indicating that the rate-determining step at least partially involves hydrogen or electron abstraction from the substrate followed by oxygen rebound or a concerted oxygenation via formation of a sigma complex between the substrate and the FeO³⁺ complex.⁸ Chemical reactivity,

therefore, is an important determinant for the SOM of a substrate. On the other hand, due to the different shapes and sizes of substrate binding cavities of the CYP enzymes and their characteristic substrate recognition features, substrate exposure to the catalytic moiety may be restricted. As a result, the most reactive site of a substrate may not be the observed SOM. This underscores the necessity of considering substrate-receptor recognition in predicting SOM.

In principle, a promising method for predicting substrate accessibility to the CYP catalytic moiety is docking simulation. Many docking studies aimed at predicting substrate sites of metabolism have been published in recent years.^{9–12} Major challenges for this approach include difficulties associated with proper representation of a protein and proper accounting of its flexibility, the lack of a score function that does not require tuning by a user with *a priori* knowledge for ranking the docked poses, and insufficient force field parameters to describe interactions between the substrate and FeO³⁺ complex.¹³ Recently, Moors et al. demonstrated an approach to account for protein flexibility in a CYP2D6 docking study.¹⁴ They

Received: March 20, 2012

Published: May 25, 2012

Report Documentation Page		<i>Form Approved OMB No. 0704-0188</i>
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.		
1. REPORT DATE 25 MAY 2012	2. REPORT TYPE	3. DATES COVERED 00-00-2012 to 00-00-2012
4. TITLE AND SUBTITLE 2D SMARTCyp Reactivity-Based Site of Metabolism Prediction for Major Drug-Metabolizing Cytochrome P450 Enzymes		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command,Biotechnology High Performance Computing Software Applications Institute,Telemedicine and Advanced Technology Research Center,Fort Detrick,MD,21702		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited		
13. SUPPLEMENTARY NOTES		

14. ABSTRACT

Cytochrome P450 (CYP) 3A4, 2D6, 2C9, 2C19, and 1A2 are the most important drug-metabolizing enzymes in the human liver. Knowledge of which parts of a drug molecule are subject to metabolic reactions catalyzed by these enzymes is crucial for rational drug design to mitigate ADME/toxicity issues. SMARTCyp, a recently developed 2D ligand structure-based method, is able to predict site-specific metabolic reactivity of CYP3A4 and CYP2D6 substrates with an accuracy that rivals the best and more computationally demanding 3D structure-based methods. In this article, the SMARTCyp approach was extended to predict the metabolic hotspots for CYP2C9, CYP2C19, and CYP1A2 substrates. This was accomplished by taking into account the impact of a key substrate-receptor recognition feature of each enzyme as a correction term to the SMARTCyp reactivity. The corrected reactivity was then used to rank order the likely sites of CYP-mediated metabolic reactions. For 60 CYP1A2 substrates, the observed major sites of CYP1A2 catalyzed metabolic reactions were among the top-ranked 1, 2, and 3 positions in 67%, 80%, and 83% of the cases respectively. The results were similar to those obtained by MetaSite and the reactivity + docking approach. For 70 CYP2C9 substrates, the observed sites of CYP2C9 metabolism were among the top-ranked 1, 2, and 3 positions in 66%, 86%, and 87% of the cases, respectively. These results were better than the corresponding results of StarDrop version 5.0, which were 61%, 73% and 77%, respectively. For 36 compounds metabolized by CYP2C19, the observed sites of metabolism were found to be among the top-ranked 1, 2, and 3 sites in 78%, 89%, and 94% of the cases, respectively. The computational procedure was implemented as an extension to the program SMARTCyp 2.0. With the extension, the program can now predict the site of metabolism for all five major drug-metabolizing enzymes with an accuracy similar to or better than that achieved by the best 3D structure-based methods. Both the Java source code and the binary executable of the program are freely available to interested users.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified	Same as Report (SAR)	15	

generated an ensemble of 1,000 CYP2D6 conformations starting with the X-ray structure of apo-CYP2D6. Docking simulations were performed on every conformer of the ensemble for many CYP2D6 substrates. When information from the docking simulations was combined with the estimated site reactivity of the substrates, reliable SOM predictions were achieved. However, even with the availability of today's massively parallel computers, docking simulations using an ensemble of thousands of protein conformations and processing a huge number of docked poses are still time-consuming and not practical for virtual screening of a large number of compounds.

Moors and co-workers also showed that predictions based on docking to any single protein conformation are significantly less reliable as judged by the scores of receiver operating characteristic (ROC) curves of their predictions.¹⁴ This is in agreement with results of a recent study by Afzelius et al. who evaluated SOM predictions for CYP3A4 and CYP2C9 based on docking using Dock and Glide software.¹⁵ For some structurally diverse CYP2C9 and CYP3A4 substrates, they found that the top-ranked sites, based on docking to a single protein conformation, were among the observed SOM only 30% to 40% of the time. This accuracy is lower than predictions given by an experienced biotransformation scientist, which were approximately 50%.¹⁵ Furthermore, it is significantly lower than predictions based on C–H bond orders of the substrates, calculated by density functional theory (B3LYP) with a small 3-21G basis set, which achieved around 60% accuracy.¹⁵

Chemical reactivity is the basis of SOM predictions of CYP3A4-mediated metabolism by Singh et al. who estimated hydrogen abstraction energies by AM1 molecular orbital calculations and predicted SOM by the energies and surface area of the hydrogen atoms.¹⁶ AM1 estimation of substrate reactivity is also the basis of SOM prediction by StarDrop, a commercial software package, which makes on-the-fly AM1 calculations.¹⁷ In addition, StarDrop makes corrections to the AM1 energies to account for steric accessibility and orientation effects via models trained using a large number of substrates of CYP3A4, 2D6, and 2C9, the top three major drug-metabolizing P450 enzymes.

Another popular commercial package for SOM prediction is MetaSite.¹⁸ It identifies likely sites of metabolism by fitness of 3D structures of a substrate oriented within the catalytic sites of the CYP enzymes represented by GRID molecular interaction fields.¹⁹ For a large number of CYP3A4 substrates, Zhou et al. found that the observed SOM were among the three top-ranked sites by MetaSite in 78% of the cases studied.²⁰ However, the success rate for the observed SOM to be ranked the highest is significantly lower.¹⁵

Generally speaking, 3D molecular structure-based prediction methods, such as docking into a large number of protein conformations and/or quantum mechanical calculation of substrate reactivity at appropriate levels of theory, are computationally demanding and not easily applicable to virtual screening of a large number of compounds. An alternative is to estimate the reactivity of molecular fragments by high-level quantum mechanical calculations on representative molecules and assign reactivity to different sites of a substrate by matching structural patterns. This is the approach used by SMARTCyp, a 2D substrate structure-based SOM prediction method.²¹ To predict SOM of CYP3A4-mediated reactions, the SMARTCyp energy of each potential site is corrected by an accessibility descriptor. The corrected energies are then used to rank order

likely sites of CYP3A4 catalyzed reactions. For 394 CYP3A4 substrates, the observed SOM were among the top-ranked one, two, and three sites 65%, 76%, and 81% of the time, respectively. They are at least the same or better than the performance of StarDrop²¹ and MetaSite.²² To predict CYP2D6 catalyzed SOM, SMARTCyp introduced two 2D descriptors to correct the energy of each atom, an accessibility descriptor and a descriptor to account for the impact of the characteristic receptor – ligand recognition between CYP2D6 and a positively charged ligand.²³ When the approach was applied to predict the SOM of a large number of CYP2D6 substrates, its performance was shown to be better than available commercial packages.²³

Encouraged by the success of SMARTCyp, we wondered if the same approach could be applied to predict the SOM of other major drug-metabolizing enzymes, namely CYP1A2, CYP2C9, and CYP2C19. In this article, we demonstrate that taking into account the effect of a single key receptor-substrate recognition feature for each of the enzymes, the SMARTCyp approach can be extended to provide accurate predictions of the SOM catalyzed by these enzymes.

2. METHOD AND RESULTS

2.1. SMARTCyp Score Functions for Ranking CYP3A4 and CYP2D6 SOM.

SMARTCyp uses the following score function to rank likely sites of CYP3A4 catalyzed reactions

$$\text{Score_3A4} = E - 8^*A(\text{kJ/mol}) \quad (1)$$

In this equation, E is an estimate of the activation energy required for a CYP-catalyzed reaction at a specific atom of the substrate. It is assigned to each atom of a substrate by matching SMARTS patterns²⁴ to a lookup table of energies derived from density functional theory calculations. The accessibility descriptor, A , is defined as the ratio between the longest bond path from a given atom divided by the longest bond path present in the whole molecule.

The SMARTCyp score function for ranking likely sites of CYP2D6-mediated metabolism is

$$\begin{aligned} \text{Score_2D6} = & E + \text{Span2End_correction} + N^*\text{Dist_correct} \\ & \text{on} \end{aligned} \quad (2)$$

where $N^*\text{Dist_correction} = 6.7 \times (8 - N^*\text{Dist})$ when $N^*\text{Dist} < 8$, and $N^*\text{Dist_correction} = 0$ when $N^*\text{Dist} \geq 8$; $\text{Span2End_correction} = 6.7 \times \text{Span2End}$ when $\text{Span2End} < 4$, and $\text{Span2End_correction} = 6.7 \times 4 + 0.01 \times \text{Span2End}$ when $\text{Span2End} \geq 4$.

The Span2End descriptor accounts for the observation that atoms in the middle of a compound are less likely to be CYP2D6-mediated SOM than atoms at the ends of the molecule. $N^*\text{Dist}$ is the maximum distance in number of chemical bonds between a likely SOM to a protonated nitrogen atom. It accounts for the fact that the Glu216 and/or Asp301 residues of CYP2D6 tend to form a characteristic salt bridge with a protonated nitrogen atom in the substrate.²⁵ Because the Glu216 and Asp301 residues are located far away from the catalytic heme moiety, substrate atoms close to the protonated nitrogen atom are less likely to be CYP2D6-mediated SOM. The constant, 6.7 kJ/mol, and the threshold values of $N^*\text{Dist}$ and Span2End were determined using a training set of 86 CYP2D6 substrates.

2.2. SOM Prediction for CYP1A2-Catalyzed Reactions.

CYP1A2 constitutes approximately 14% of human liver P450

enzymes in Caucasians²⁶ and about 18% in Asians.⁴ It is responsible for the clearance of about 5% of the top 200 drugs on the U.S. market.²⁷ Its catalytic site has been well characterized.^{28–31} The X-ray structure of the protein cocrystallized with α -naphthoflavone (ANF) indicates that the substrate-binding cavity of CYP1A2 is narrow and lined by amino acid residues that define a relatively planar substrate binding platform.²⁹ ANF is a potent and competitive inhibitor of CYP1A2-catalyzed reactions. Its high binding affinity to CYP1A2 was attributed to the overall fitness of the shape and size of the molecule to the binding site cavity and the resulting dense and extensive van der Waals interactions with the nonpolar side chains of the protein. Additionally, it was noted that there is a water molecule close to the carbonyl group of ANF that provides an extra binding interaction. This water molecule is the only one present in the active site, and there appears to be no solvent channels that connect the active site cavity with the protein surface. As shown in Figure 1, the water molecule appears to be hydrogen-bonded to the carbonyl of ANF as well as to the carbonyl of Gly316.²⁹

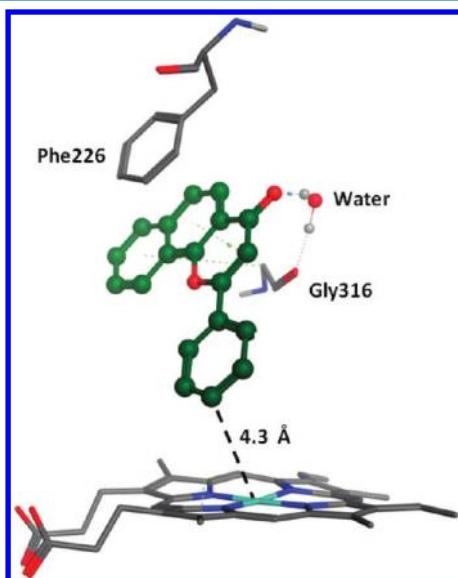


Figure 1. The X-ray crystal structure of human CYP1A2 cocrystallized with a potent inhibitor, α -naphthoflavone, showing tight binding due to $\pi-\pi$ stacking with residue Phe226 and a structured water molecule in hydrogen bonding interactions with the ligand carbonyl group at one end and with the carbonyl group of Gly316 at the other end. The ligand-protein interactions orient the ligand so that the 4' position of the phenyl ring is exposed to the heme. For clarity, other parts of the protein are not displayed.

Rydberg et al. showed that reliable SOM prediction for CYP1A2 substrates can be achieved by combining site reactivity

derived from density functional theory calculations with docking into the X-ray crystal structure of CYP1A2.³² For 60 CYP1A2 substrates, the accuracy of SOM prediction by the reactivity + docking approach is similar to that of MetaSite version 3.0, as shown in Table 1. Both the docking and MetaSite calculations require 3D molecular structure information and evaluation of the fitness of substrate molecular shape and size to the binding site cavity. Encouraged by the success of SMARTCyp for CYP3A4 and CYP2D6 SOM predictions, which use 2D molecular structure information of the substrates only, we examined whether prediction accuracies similar to that achieved by the reactivity/docking method and MetaSite could be achieved for CYP1A2 by the SMARTCyp approach.

We first examined the performance of using SMARTCyp reactivity only, without any substrate-CYP1A2 recognition information, to predict SOM of CYP1A2 mediated metabolic reactions. We examined two schemes of SMARTCyp reactivity as defined by eq 1 and by the following equation

$$\text{Score_2D6}' = E + \text{Span2End_correction} \quad (3)$$

Equation 1 is the scoring function for CYP3A4-mediated SOM. Equation 3 is equal to the scoring function for CYP2D6-catalyzed SOM minus the $N^{\star}\text{Dist_correction}$ term. The $N^{\star}\text{Dist_correction}$ term was excluded in eq 3 because Rydberg et al. used it to account for the effect of CYP2D6-substrate recognition, which is not applicable in the case for CYP1A2.

Table 1 compares the performance of SOM predictions by Score_3A4 and Score_2D6', for the 60 CYP1A2 substrates used by Rydberg et al. in their recent study,³² with the performance of MetaSite and Rydberg's reactivity + docking approach. Molecular structures of the 60 CYP1A2 substrates, observed SOM, and predicted top three SOM by Score_3A4 and Score_2D6' are provided in Figure S1 in the Supporting Information. Compared to the percentage of observed SOM in randomly picked sites (column random in Table 1), both Score_3A4 and Score_2D6' performed reasonably well, considering none of the information was CYP1A2-specific. These results indicate that chemical reactivity is the most important determinant for SOM of CYP1A2-catalyzed metabolic reactions.

Table 1 also shows that Score_2D6' outperformed Score_3A4 in all four performance criteria (percentage of the observed major sites of metabolism found in the top-ranked 1, 2, and 3 sites, as well as the percentage of observed major or minor sites of metabolism found in the highest ranked sites). It clearly indicates that Score_2D6' is a better representation of CYP1A2 substrate site reactivity than Score_3A4. This can be rationalized by noting that the CYP3A4 substrate binding cavity is large and very flexible, while the CYP1A2 binding site cavity is much smaller and narrower, and, hence, the latter is more similar to the substrate binding site cavity of CYP2D6.

Table 1. Performance Comparison for CYP1A2-Catalyzed Site of Metabolism (SOM) Prediction on 60 CYP1A2 Substrates^a

	Score_3A4 ^b	Score_2D6 ^c	Score_2D6'	Score_1A2 ^d	reactivity + docking ^e	MetaSite ^e	random
number of compounds	60	60	32 w/o rCO	28 w/rCO	60	60	60
1st ranked site is a major SOM	55	57	72	39	67	67	15
1st or 2nd ranked site is a major SOM	70	73	84	61	80	72	29
1st, 2nd, or 3rd ranked is a major SOM	78	82	91	71	83	87	42
1st ranked is a major or minor SOM	67	72	81	61	78	77	19

^aNumerical values in the table are percentages of the substrates for which the SOM were correctly ranked. ^bScore_3A4 is defined by eq 1.

^cScore_2D6' is defined by eq 3. ^dScore_1A2 is defined by eq 5. ^eReference 32.

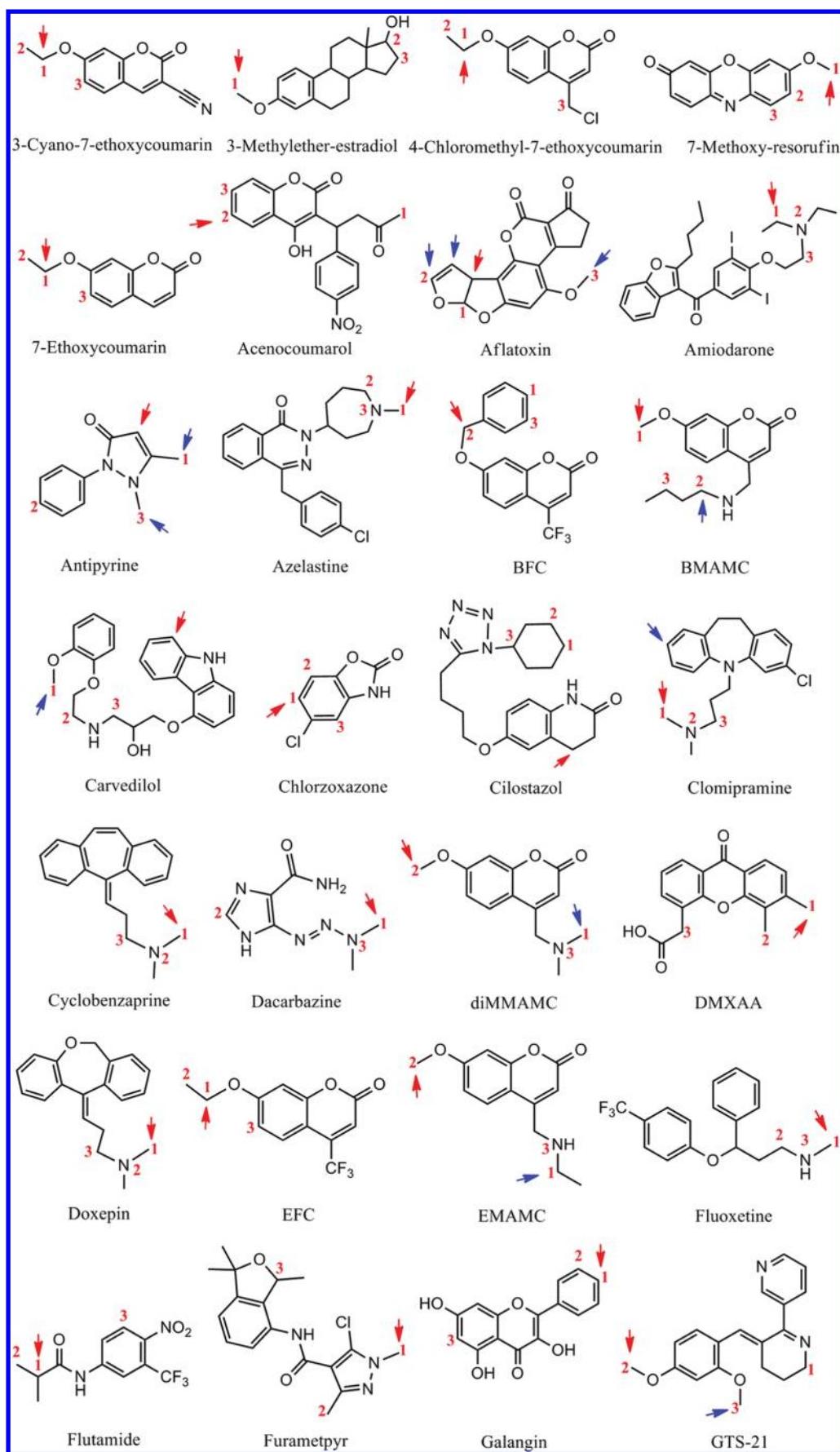


Figure 2. continued

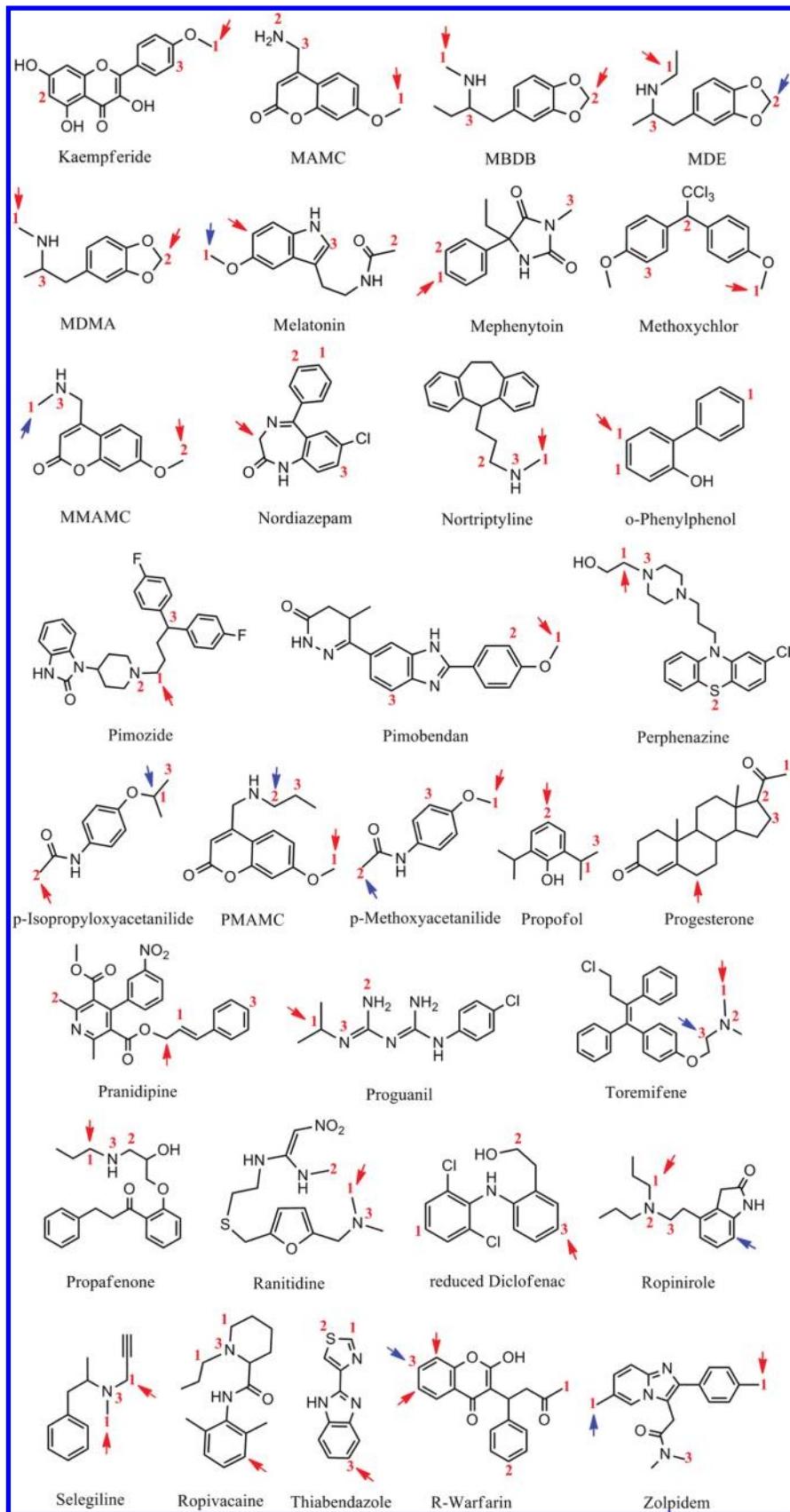


Figure 2. SOM of CYP1A2-catalyzed reactions. Red arrows: observed major sites. Blue arrows: observed minor sites. Red numbers: top-ranked sites by Score_1A2. For molecules with symmetrically equivalent sites, only one of the equivalent sites is labeled.

However, even with Score_2D6', the performance of SOM prediction for CYP1A2 substrates was generally inferior to that

of MetaSite or the reactivity + docking approach. Detailed examination of the structures of the 60 substrates revealed that

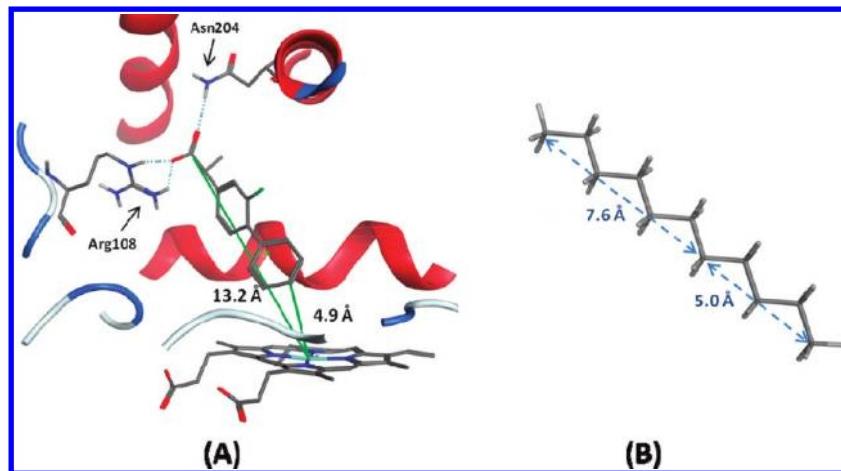


Figure 3. (A) X-ray structure of human CYP2C9 cocrystallized with flurbiprofen showing hydrogen bonding interactions between the anionic carboxyl group with the Asn204 and Arg108 side chains of the protein. For clarity, other parts of the protein are not displayed. (B) Molecular structure of n-undecane optimized by the MMFF force field.

28 of them have a carbonyl group with the carbon atom being part of a ring. For the 32 molecules without a cyclic carbonyl moiety, SOM prediction given by Score_2D6' significantly outperformed that of MetaSite and the reactivity + docking approach, as shown in Table 1. However, for the 28 compounds with a cyclic carbonyl moiety, SOM prediction given by Score_2D6' alone was significantly worse. This indicates that the cyclic carbonyl group may be an important contributor to the SOM of CYP1A2-catalyzed reactions. It could be a key substrate recognition feature of the CYP1A2 enzyme. The substrate-enzyme recognition feature may orient the substrate so that some reactive sites of the substrate are out of reach by the catalytic heme moiety. Figure 1 shows a likely interaction between the carbonyl group of a substrate and the enzyme - the hydrogen bonding interaction between the carbonyl oxygen and a water molecule. The water molecule, in turn, forms hydrogen bonding interactions with the carbonyl oxygen of Gly316 of the protein, as reported by Sansen et al.²⁹ This is an example of a water molecule serving as a bridge for receptor-ligand recognition. The fact that a ring carbonyl is required for this receptor-ligand interaction can be explained by the narrow and flat binding cavity of the enzyme. A noncyclic carbonyl group may not be as effectively anchored to the binding site because of steric hindrance from the left and right groups attached to the carbonyl carbon atom. The ring ties the two groups back, making the carbonyl oxygen effectively exposed for hydrogen bonding interactions with the water molecule. The ring moiety fits the flat binding cavity well and forms favorable van der Waals interactions with the hydrophobic side chains of the protein.

Figure 1 shows that, because of the ring carbonyl-water hydrogen bonding interactions, carbon atoms at the 3', 4', and 5' positions of the phenyl ring are about 5 Å from the heme iron, a distance generally considered to be within the reach of the FeO³⁺ catalytic moiety of the heme.³³ On the other hand, atoms within four chemical bonds of the carbonyl carbon atom appear too far from the catalytic moiety. However, this is only a static picture of a protein-bound ligand. In reality, both the protein and the ligand have some degrees of freedom. In light of the SMARTCyp approach for predicting SOM of CYP2D6-mediated metabolic reactions, we introduced a correction term to account for the effect of substrate-CYP1A2 recognition. This term is called *Dist2rCO_correction* and is defined as

$$\text{Dist2rCO_correction} = c \times (\text{Dist2rCO_cutoff} - \text{Dist2rCO}) \quad (4)$$

In eq 4, *Dist2rCO* denotes the distance in number of chemical bonds between a substrate atom of interest and the most distant cyclic carbonyl carbon of the substrate. The *Dist2rCO_cutoff* is a threshold bond distance above which no correction is needed. Equation 4 is similar to the *N^i Dist_correction* term in eq 2 for CYP2D6-catalyzed SOM. Rydberg and Olsen selected the constant *c* to be 6.7 kJ/mol and *N^i Dist_cutoff* to be 8 for CYP2D6 by using a training set of 86 CYP2D6 substrates. In the case of CYP1A2, we do not have a large number of substrates as a training set for determining these constants. However, we found that the results were not very sensitive to the value of the constant *c*, as long as it was approximately 10 kJ/mol. Therefore, for simplicity, we assigned *c* = 10 kJ/mol and experimented with bond distance cutoff values of *Dist2rCO_cutoff*. For the 28 substrates with cyclic carbonyl groups, the best performance was achieved with *Dist2rCO_cutoff* = 6. In the end, our score function for ranking likely SOM for CYP1A2-catalyzed metabolic reactions was

$$\text{Score_1A2} = E + \text{Span2End_correction} + \text{Dist2rCO_correc} \\ \text{tion} \quad (5)$$

where *Dist2rCO_correction* = $10.0 \times (6 - \text{Dist2rCO})$ when *Dist2rCO* < 6, and *Dist2rCO_correction* = 0 when *Dist2rCO* ≥ 6.

With the *Dist2rCO_correction* term to account for the effect of CYP1A2 - substrate recognition, the SOM prediction based on Score_1A2 improved significantly. Performance of SOM prediction for the 60 substrates is shown in Table 1 under the column Score_1A2. The molecular structures of the 60 substrates, observed SOM, and predicted top three SOM by Score_1A2 are given in Figure 2.

Based on all four criteria in Table 1, the 2D ligand-based Score_1A2 is at least as good as the 3D structure-based MetaSite and the reactivity + docking approach, indicating that the water mediated substrate recognition feature is an important determinant for CYP1A2-catalyzed metabolic reactions.

2.3. SOM Prediction for CYP2C9-Catalyzed Reactions. According to Rowland-Yeo et al., nearly 20% of the CYP enzymes in the human liver are CYP2C9.²⁶ CYP2C9

metabolizes about 15% of marketed drugs²⁷ and is known to exhibit selectivity for the oxidation of relatively small, lipophilic anions.³⁴ Figure 3 shows the X-ray crystal structure of human CYP2C9 cocrystallized with flurbiprofen. The structure indicates that the carboxyl group of flurbiprofen forms hydrogen bonding interactions with the Arg108 and Asn204 side chains of the protein.³⁵ Since the Arg108 and Asn204 side chains are at the opposite end to the heme in the substrate binding cavity, substrate atoms in close proximity to the carboxyl group that forms hydrogen bonding interactions with the side chains are less likely to be the SOM because of their distance from the catalytic moiety. The high resolution crystal structure showed that the carboxyl carbon atom of flurbiprofen is 13.2 Å from the heme iron. The distance is close to 10 C–C single bonds (~12.6 Å), as shown in Figure 3. It is generally considered that substrate atoms within ~5 Å of the heme iron are accessible by the FeO³⁺ catalytic moiety and may become the site of CYP-mediated metabolic reactions.³³ Based on the MMFF force field calculations, 5 Å is approximately the distance between the terminal carbon atoms in n-pentane, or four C–C bonds, as shown in Figure 3(B). Taken together, Figure 3 indicates that substrate atoms within five chemical bonds of the carboxyl group have reduced probability to be sites of CYP2C9 catalyzed metabolic reactions. Furthermore, the closer an atom is to the carboxyl group, the less likely it is to be metabolized by CYP2C9. On the basis of this key substrate-receptor recognition feature and consistent with the SMART-Cyp approach, a reasonable score function for ranking likely sites of CYP2C9 metabolism is

$$\text{Score_2C9} = E + \text{Span2End_correction} + \text{Dist2CO_correct}_{\text{ion}} \quad (6)$$

where $\text{Dist2CO_correction} = 10 \times (6 - \text{Dist2CO})$ when $\text{Dist2CO} \leq 5$, and $\text{Dist2CO_correction} = 0$ when $\text{Dist2CO} \geq 6$.

In eq 6, Dist2CO is the distance in number of chemical bonds between a substrate atom of interest and the most distant carboxyl carbon of the substrate. The $\text{Dist2CO_correction}$ term accounts for the effect of substrate - receptor recognition on CYP2C9 SOM.

To test this scoring function, we applied it to predict the SOM of 21 carboxylic acids collated by Sykes et al.³⁶ that are metabolized by CYP2C9. Table 2 gives the performance of the

Table 2. Performance Comparison for CYP2C9-Catalyzed Site of Metabolism (SOM) Predictions on 70 CYP2C9 Substrates^a

	Score_2C9 ^b			StarDrop 5.0 ^c		
	carboxylic acids	non-acid	overall	carboxylic acids	non-acid	overall
1st ranked site is the observed SOM	67	65	66	57	63	61
1st or 2nd ranked site is the observed SOM	81	88	86	71	73	73
1st, 2nd, or 3rd ranked site is the observed SOM	86	88	87	71	79	77

^aNumerical values in the table are percentages of the substrates for which the observed SOM were correctly ranked. ^bScore_2C9 is defined eq 6. ^cPrediction given by StarDrop version 5.0 on 69 of the 70 substrates. StarDrop calculation on one of the 70 compounds failed.

score function as measured by the percentage of cases of the experimentally observed SOM were among the top-ranked 1, 2, and 3 sites. For comparison, the performance of SOM prediction by StarDrop version 5.0 is also included in Table 2. By all three performance measures given in Table 2, the score function of eq 6 outperformed StarDrop 5.0 for the 21 carboxylic acid substrates.

To evaluate if it is necessary to apply the progressive energy penalty for atoms in close proximity to the carboxyl group, we also made SOM predictions using Score_2D6' only. The corresponding values of the correctly predicted SOM for the 21 carboxylic acids are 52%, 67%, and 67%, respectively. The significantly inferior results obtained by using Score_2D6' underscores the importance of the key substrate - receptor recognition in determining the regioselectivity of CYP2C9-catalyzed metabolic reactions. These results also validate the use of *Dist2CO_correction* to account for effects of this substrate - receptor recognition.

We also applied Score_2D6' to rank likely sites of CYP2C9 metabolism for the 49 noncarboxylic substrates collated by Sykes et al. For these compounds, the percentages of the observed SOM being among the top-ranked 1, 2, and 3 sites were 57%, 76%, and 80%, respectively. The results are close to the performance of StarDrop 5.0 for these compounds but significantly worse than the performance of Score_2C9 for the carboxylic acids. Inspection of the molecular structures of the compounds revealed that, even though they are noncarboxylic acids, many of them have carbonyl groups. Carbonyl groups are perhaps weaker hydrogen bond acceptors than the negatively charged carboxyl groups, but nonetheless, they are hydrogen bond acceptors. There is no reason to believe that they do not form hydrogen bonding interactions with the Arg108 or Asn204 side chains of the CYP2C9 enzyme. Failing to account for this hydrogen bonding interaction might be responsible for the inferior performance of the score function for noncarboxylic substrates. To test this hypothesis, we redefined *Dist2CO* in eq 6. For a carboxylic acid or anion, it does not matter if there are other carbonyl groups or not, *Dist2CO* is always defined as the distance in number of chemical bonds between a substrate atom of interest and the most distant carboxylic carbon atom of the substrate. For other substrates, *Dist2CO* is defined as the distance in number of chemical bonds between a substrate atom of interest and the most distant carbonyl carbon of the substrate. Note that for a compound with both carboxyl and carbonyl groups, the carboxyl group is given precedence in calculating *Dist2CO*. This is based on the observation that the carboxyl group in Figure 3 forms hydrogen bonding interactions with both the Arg108 and Asn204 residues of the protein. A noncarboxyl carbonyl group most likely forms weaker hydrogen bonding interactions with only one of the two protein residues. This is consistent with the observation that CYP2C9 exhibits selectivity for the oxidation of carboxylic acid substrates. With this modification, Score_2C9 was applied to rank likely SOM of the 49 noncarboxylic acid substrates. Table 2 shows that the results were significantly improved. Overall, for the 70 CYP2C9 substrates, the observed SOM were found in the top-ranked 1, 2, and 3 sites in 66%, 86%, and 87% of the cases, respectively. For comparison, StarDrop 5.0 was also used to predict SOM for the 70 compounds. However, our StarDrop calculations were successful for only 69 of the 70 compounds. The calculation for zafirlukast failed, presumably due to some problem in the semiempirical AM1 stage. The corresponding percentages for the 69 compounds achieved by StarDrop were

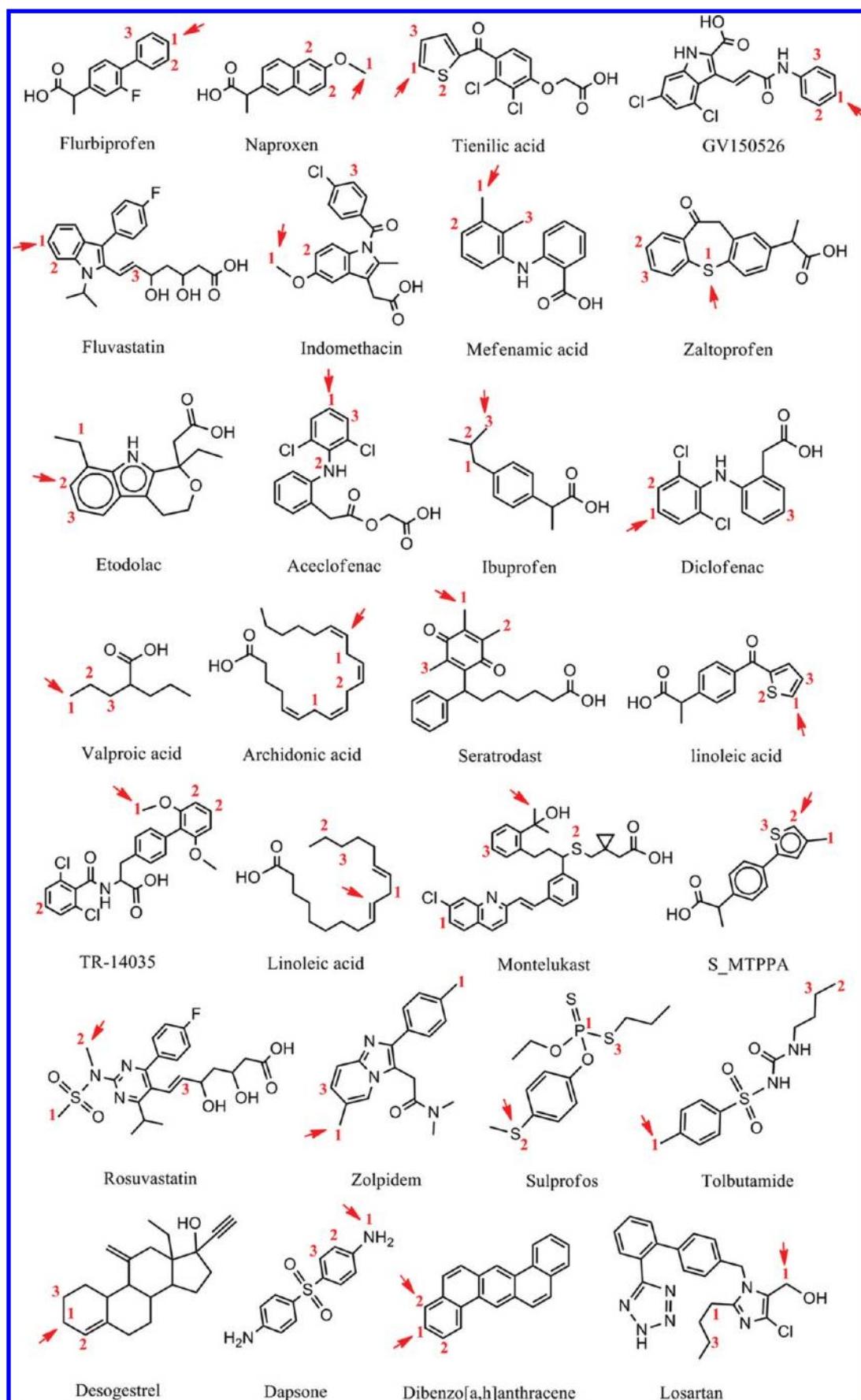


Figure 4. continued

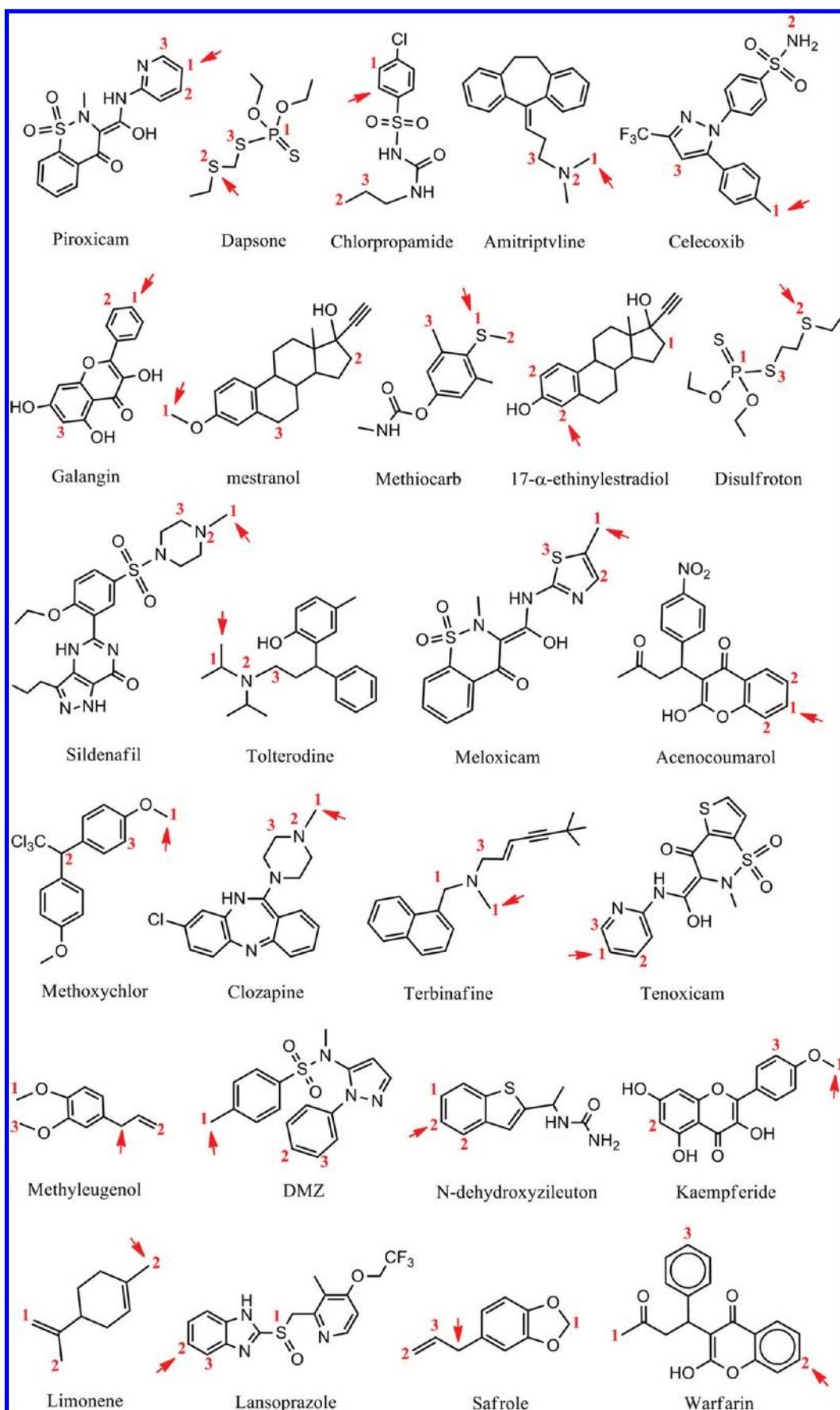


Figure 4. continued

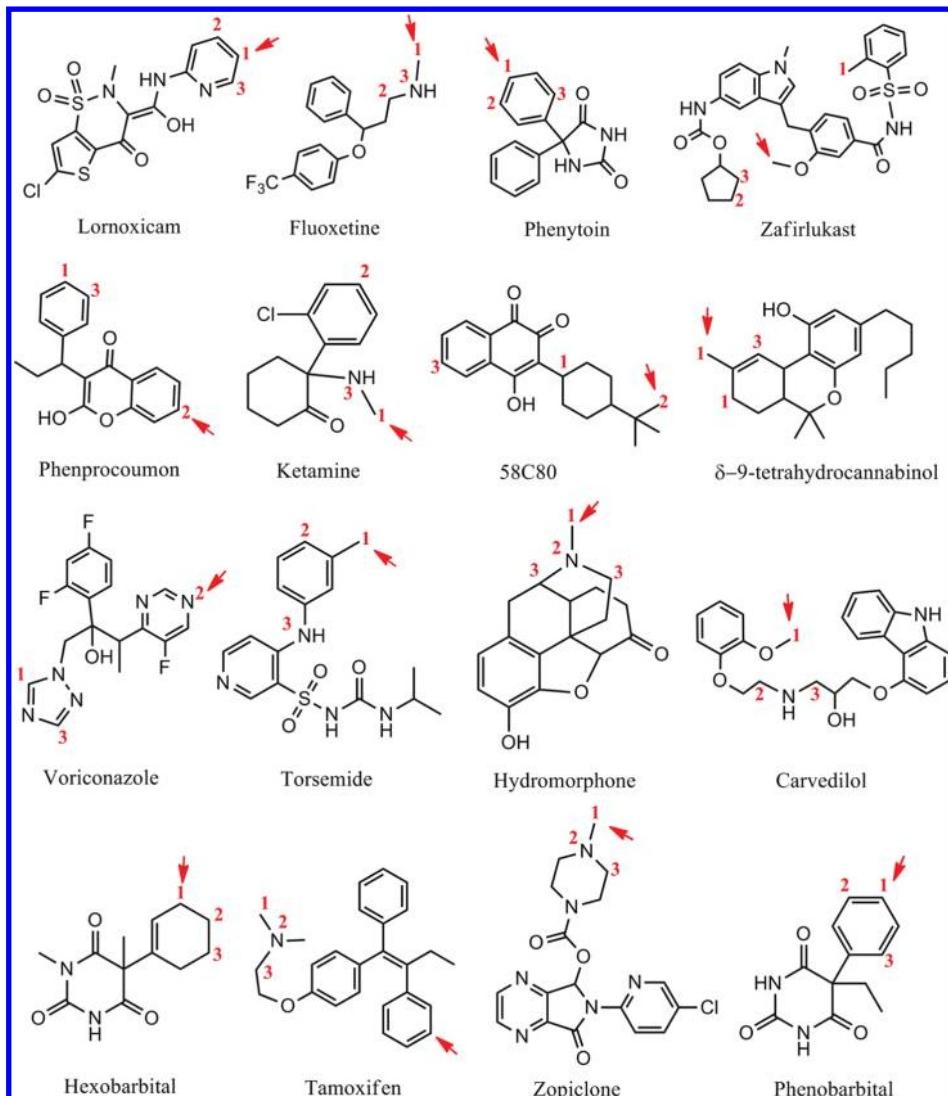


Figure 4. SOM of CYP2C9-catalyzed reactions. Red arrows: observed sites of metabolism. Red numbers: top-ranked sites by Score_2C9. For molecules with symmetrically equivalent sites, only one of the equivalent sites is labeled.

61%, 73%, and 77%, respectively. Figure 4 shows molecular structures of the 70 CYP2C9 substrates, the experimentally observed SOM, and the three top-ranked sites by Score_2C9. The fact that the redefined *Dist2CO_correction* term improved SOM prediction supports the hypothesis that carbonyl groups also form hydrogen bonding interactions with CYP2C9, which influences the SOM of CYP2C9-catalyzed metabolic reactions.

2.4. SOM Prediction for CYP2C19-Catalyzed Metabolic Reactions. CYP2C9 and CYP2C19 have a very high level of sequence identity. The two proteins have only 43 differing amino acid residues among a total sequence length of 490 residues.³⁷ However, they have quite different substrate specificity. For instance, while CYP2C9 shows selectivity for and is a major metabolizer of acidic substrates that are anionic under physiological pH, most signature CYP2C19 substrates are lipophilic and neutral at physiological pH.³⁸ In addition, CYP2C19 is highly selective for 4'-hydroxylation of mephenytoin and 5-hydroxylation of proton pump inhibitors such as omeprazole and lansoprazole, while CYP2C9 has little activity for these compounds.³⁹ Because of this, mephenytoin, omeprazole, and lansoprazole were termed marker substrates of CYP2C19 by Wada et al.³⁹

Even though a crystal structure of human CYP2C19 is not yet available, mutation studies have identified some key amino acids that are crucial for conferring CYP2C9 and CYP2C19 substrate specificity.⁴⁰ For example, molecular structural analysis indicated that the F-G loop forms a flexible lid and a substrate entrance channel in CYP2C8, CYP2C9, and CYP2B4.^{41–43} Substitution of the F-G loop in CYP2C9 to that of CYP2C19, i.e., Ser220 → Pro and Pro221 → Thr, does not alter the enzyme activity toward CYP2C9 marker substrates but enhanced 4'-hydroxylation of mephenytoin and 5-hydroxylation of omeprazole, which were not detectable in CYP2C9.³⁹ In addition, mutation of Ile99 to His99 (of CYP2C19) in CYP2C9 increased omeprazole 5-hydroxylase to ~51% of that of CYP2C19.⁴⁰ This provides strong evidence that amino acids 99, 220, and 221 are among the key residues that determine the distinctive substrate specificities of CYP2C9 and CYP2C19.

The crystal structure of CYP2C9 indicates that Ile99 is very close to the heme,³⁵ and it is expected that this is also the case for His99 in CYP2C19. Locuson et al. postulated that His99 plays an important role in CYP2C19 metabolism by serving as a hydrogen bond donor and forming hydrogen bonding

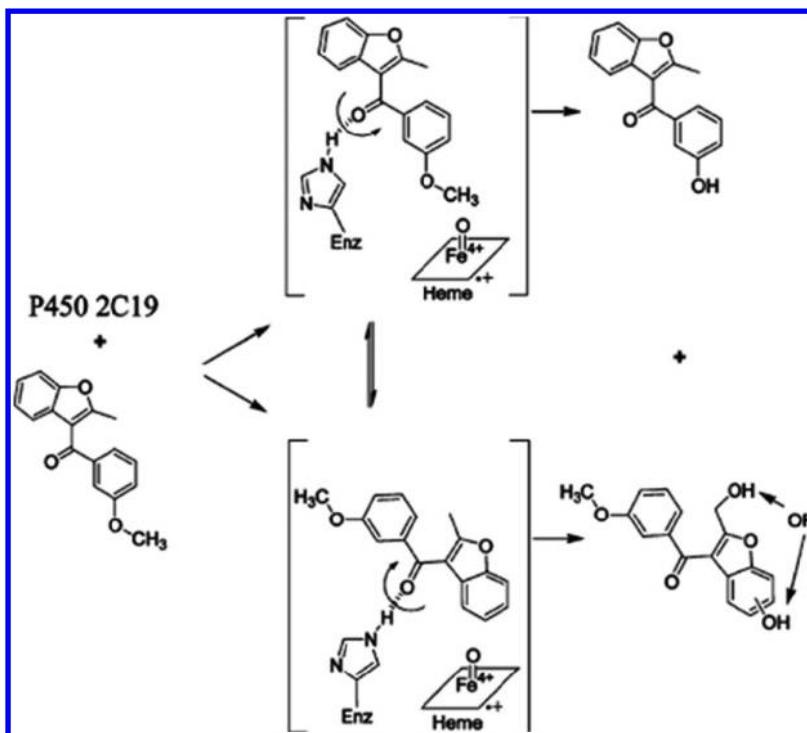


Figure 5. Hydrogen bonding interactions, postulated by Locuson et al., between His99 of CYP2C19 and a substrate, and the role of the hydrogen bonding interactions in the regioselectivity of the metabolic reactions. The sketch was adapted with permission from the paper of Locuson et al. published in *J. Med. Chem.* 2004, 47, 6768–6776 Copyright 2004, American Chemical Society.

interactions with the carbonyl or sulfinyl oxygen of the substrate, as illustrated in Figure 5.⁴⁴ The hydrogen bonding interactions enhance the affinity of the substrates to the enzyme and orient the substrates to the nearby heme catalytic site.

To examine if the SMARTCyp approach could be extended to predict SOM of CYP2C19 mediated metabolic reactions, we collated 36 compounds metabolized by CYP2C19. The observed CYP2C19 SOM of these compounds were obtained from the papers of Lewis et al.³⁸ and Locuson et al.⁴⁴ and from DrugBank.⁴⁵

As a first step, we applied Score_2D6' to evaluate if SMARTCyp reactivity alone was sufficient for predicting SOM of the 36 CYP2C19 substrates. As shown in Figure 6 and Table 3, reasonably reliable predictions were achieved by Score_2D6' without using any CYP2C19-specific information. The observed sites of CYP2C19 metabolism were found to be among the top-ranked 1, 2, and 3 sites in 69%, 89%, and 92% of the cases, respectively. However, for the CYP2C19 marker substrates mephenytoin and lansoprazole, the observed sites of metabolism were second-ranked sites by Score_2D6', as shown in Figure 6. To improve prediction consistent with the SMARTCyp approach, we adopted the following score function

$$\text{Score_2C19} = E + \text{Span2End_correction} + \text{Dist2XO_correction} \quad (7)$$

$\text{Dist2XO_correction} = 10 \times (\text{Cutoff} - \text{Dist2XO})$, when $\text{Dist2XO} < \text{Cutoff}$, and $\text{Dist2XO_correction} = 0$ otherwise.

In the preceding equations, XO represents $\text{C}=\text{O}$ or $\text{S}=\text{O}$ moieties, as they were postulated by Locuson et al.⁴⁴ to be CYP2C19 substrate recognition features. Dist2XO denotes the distance in number of chemical bonds between a substrate atom of interest to the most distant $\text{C}=\text{O}$ or $\text{S}=\text{O}$ group of the substrate.

Test calculations indicated that a *Cutoff* value of 4 is reasonable. This threshold value implies that substrate atoms within three chemical bonds of $\text{C}=\text{O}$ or $\text{S}=\text{O}$ are likely to be positioned too far from the catalytic heme moiety and, therefore, have lower probabilities to be sites of CYP2C19-catalyzed reactions. Figure 6 shows the top three sites for each substrate ranked by Score_2D6' and Score_2C19. Overall, an observed SOM was found to be among the top 1, 2, and 3 sites ranked by Score_2C19 78%, 89%, and 94% of the time, respectively, for the 36 substrates. Two factors contributed to the higher percentage of correct SOM predictions for CYP2C19 substrates than for CYP1A2 and CYP2C9 substrates. First, the number of CYP2C19 substrates we collected and used in the study was relatively small. Second, and more importantly, the observed metabolic reactions for approximately half of the CYP2C19 substrates were either N- or O-dealkylation reactions. These reactions have significantly lower activation barriers than most other reactions. As a result, the SOM of N- and O-dealkylation reactions were correctly predicted by SMARTCyp energies for most of the compounds.

3. CONCLUSIONS

The results of this study, together with those of Rydberg et al. on CYP3A4 and CYP2D6 SOM predictions, demonstrated that chemical reactivity is the most important determinant of substrate SOM of CYP-catalyzed reactions. They also demonstrated that the activation energies of Rydberg et al., derived from density functional theory calculations and corrected by the Span2End descriptor, are reasonable representations of substrate site reactivity for most CYP enzymes. Highly reliable SOM predictions for CYP1A2, CYP2C9, and CYP2C19 were derived based on the reactivity combined with a correction term to account for the effect of key substrate - enzyme recognition on substrate access to the catalytic moiety. The fact

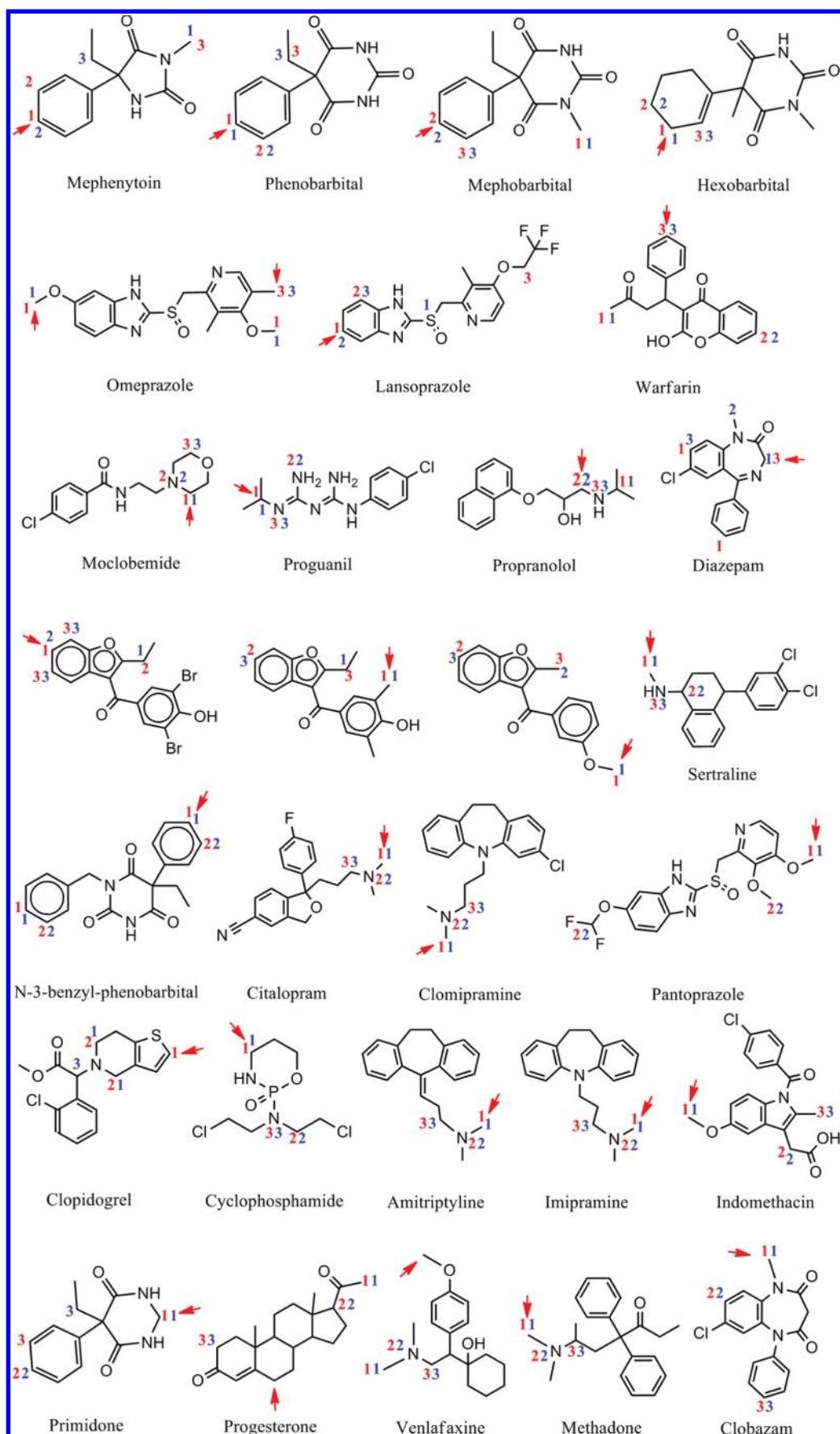


Figure 6. continued

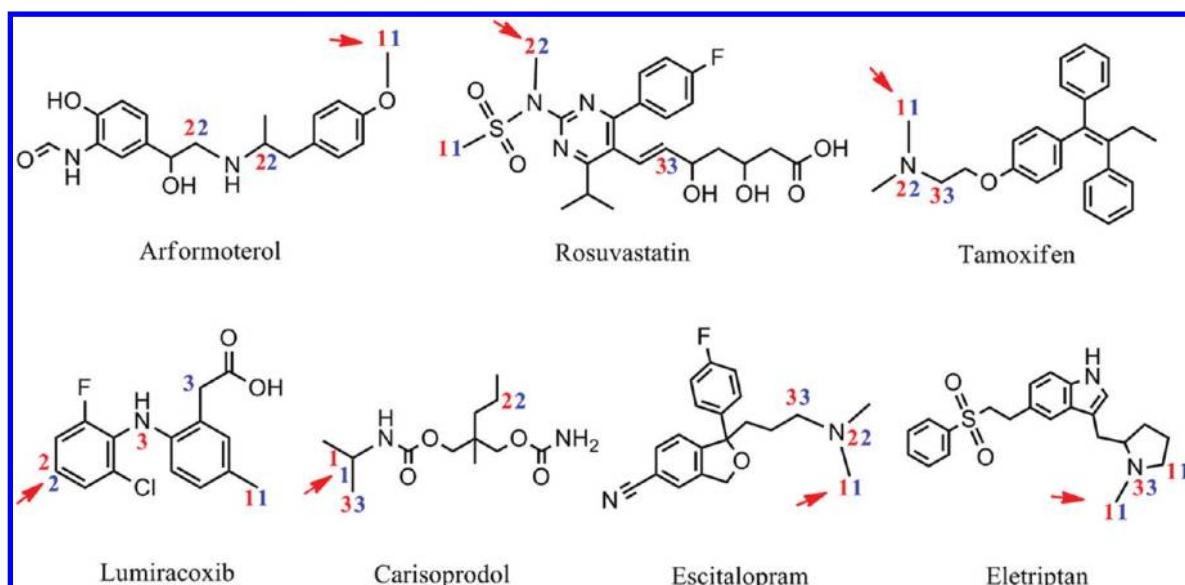


Figure 6. SOM of CYP2C19-catalyzed reactions. Red arrows: observed sites of metabolic reactions. Blue numbers: top-ranked sites by Score_2D6' (no CYP2C19 specific information is used). Red numbers: top-ranked sites by Score_2C19. For molecules with symmetrically equivalent sites, only one of the equivalent sites is labeled.

Table 3. Performance Comparison for CYP2C19-Catalyzed Site of Metabolism (SOM) Predictions on 36 CYP2C19 Substrates^a

	Score_2D6 ^b	Score_2C19 ^c
1st ranked site is the observed SOM	69	78
1st or 2nd ranked site is the observed SOM	89	89
1st, 2nd, or 3rd ranked site is the observed SOM	92	94

^aNumerical values in the table are percentages of the substrates for which the observed SOM were correctly ranked. ^bScore_2D6' is defined by eq 3. ^cScore_2C19 is defined by eq 7.

that the correction term defined by the distance to a cyclic carbonyl group significantly improved CYP1A2 SOM predictions supports the finding that a structured water molecule in the CYP1A2 substrate binding cavity plays an important role in CYP1A2 substrate recognition. This water molecule was observed to be hydrogen-bonded to the carbonyl of a CYP1A2 inhibitor and to the carbonyl of Gly316 of the enzyme. Similarly, the fact that a correction term, defined by the distance to a substrate carbonyl or sulfinyl group, improved CYP2C19 SOM prediction supports the hypothesis of Locuson et al. on the role of residue His99 in CYP2C19 metabolism.

In principle, docking of substrates into the protein structures should give reliable prediction of substrate sites that may be accessible to the catalytic moiety. However, docking simulations are much more computationally demanding than the 2D substrate structure based SMARTCyp approach for SOM prediction. In addition, it is challenging to properly account for protein flexibility in a docking simulation, and it requires expert knowledge to select an appropriate docking score function to rank the docked poses. As a result, docking is more of an expert tool that requires significant knowledge and experience to process computational results and select relevant poses from a large number of docked conformations. On the other hand, the SMARTCyp approach for predicting likely sites of CYP-catalyzed metabolic reactions is very fast as it uses 2D molecular structure information only and is easily applied by

anyone without the need for specialized training. The score functions for ranking likely sites of metabolism by CYP1A2, 2C9, and 2C19 enzymes as described in this article were implemented as an extension of the SMARTCyp 2.0 program by modifying the program modules. With the extension to cover the three additional CYP isoforms, the program can now successfully predict SOM for all five major drug metabolizing CYP enzymes. Both the Java source code and the binary executable of the program are freely available for download at http://www.bhsai.org/downloads/smartercyp_ext/.

ASSOCIATED CONTENT

S Supporting Information

Figure S1. SOM of CYP1A2 catalyzed reactions predicted by Score_3A4 (eq 1) and by Score_2D6' (eq 3). This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*Phone: 301-619-1979. Fax: 301-619-1983. E-mail: rliu@bhsai.org.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Funding of this research was provided by the U.S. Department of Defense Threat Reduction Agency Grant TMTI0004_09_BH_T. The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the U.S. Army or of the U.S. Department of Defense. We thank Mr. Li Chen and Jason Smith for their technical assistance. This paper has been approved for public release with unlimited distribution.

REFERENCES

- (1) Sigel, A.; Sigel, H.; Sigel, R. *The Ubiquitous Roles of Cytochrome P450 Proteins: Metal Ions in Life Sciences*; Wiley: 2007; p 3.

- (2) Danielson, P. B. The cytochrome P450 superfamily: biochemistry, evolution and drug metabolism in humans. *Curr. Drug Metab.* **2002**, *3*, 561–597.
- (3) Guengerich, F. P. Cytochrome p450 and chemical toxicology. *Chem. Res. Toxicol.* **2008**, *21*, 70–83.
- (4) Shimada, T.; Yamazaki, H.; Mimura, M.; Inui, Y.; Guengerich, F. P. Interindividual variations in human liver cytochrome P-450 enzymes involved in the oxidation of drugs, carcinogens and toxic chemicals: studies with liver microsomes of 30 Japanese and 30 Caucasians. *J. Pharmacol. Exp. Ther.* **1994**, *270*, 414–423.
- (5) Lynch, T.; Price, A. The effect of cytochrome P450 metabolism on drug response, interactions, and adverse effects. *Am. Fam. Physician* **2007**, *76*, 391–396.
- (6) Slaughter, R. L.; Edwards, D. J. Recent advances: the cytochrome P450 enzymes. *Ann. Pharmacother.* **1995**, *29*, 619–624.
- (7) Wilkinson, G. R. Drug metabolism and variability among patients in drug response. *New Engl. J. Med.* **2005**, *352*, 2211–2221.
- (8) Guengerich, F. P. Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. *Chem. Res. Toxicol.* **2001**, *14*, 611–650.
- (9) de Graaf, C.; Oostenbrink, C.; Keizers, P. H.; van der Wijst, T.; Jongejan, A.; Vermeulen, N. P. Catalytic site prediction and virtual screening of cytochrome P450 2D6 substrates by consideration of water and rescoring in automated docking. *J. Med. Chem.* **2006**, *49*, 2417–2430.
- (10) de Groot, M. J.; Ackland, M. J.; Horne, V. A.; Alex, A. A.; Jones, B. C. Novel approach to predicting P450-mediated drug metabolism: development of a combined protein and pharmacophore model for CYP2D6. *J. Med. Chem.* **1999**, *42*, 1515–1524.
- (11) Park, J. Y.; Harris, D. Construction and assessment of models of CYP2E1: predictions of metabolism from docking, molecular dynamics, and density functional theoretical calculations. *J. Med. Chem.* **2003**, *46*, 1645–1660.
- (12) Vasanthanathan, P.; Hritz, J.; Taboureau, O.; Olsen, L.; Jorgensen, F. S.; Vermeulen, N. P.; Oostenbrink, C. Virtual screening and prediction of site of metabolism for cytochrome P450 1A2 ligands. *J. Chem. Inf. Model.* **2009**, *49*, 43–52.
- (13) Yuriev, E.; Agostino, M.; Ramsland, P. A. Challenges and advances in computational docking: 2009 in review. *J. Mol. Recognit.* **2011**, *24*, 149–164.
- (14) Moors, S. L.; Vos, A. M.; Cummings, M. D.; Van Vlijmen, H.; Ceulemans, A. Structure-based site of metabolism prediction for cytochrome P450 2D6. *J. Med. Chem.* **2011**, *54*, 6098–6105.
- (15) Afzelius, L.; Arnby, C. H.; Broo, A.; Carlsson, L.; Isaksson, C.; Jurva, U.; Kjellander, B.; Kolmodin, K.; Nilsson, K.; Raubacher, F.; Weidolf, L. State-of-the-art tools for computational site of metabolism predictions: comparative analysis, mechanistical insights, and future applications. *Drug Metab. Rev.* **2007**, *39*, 61–86.
- (16) Singh, S. B.; Shen, L. Q.; Walker, M. J.; Sheridan, R. P. A model for predicting likely sites of CYP3A4-mediated metabolism on drug-like molecules. *J. Med. Chem.* **2003**, *46*, 1330–1336.
- (17) StarDrop, version 5.0; Optibrium Ltd.: Cambridge, United Kingdom, 2011.
- (18) Cruciani, G.; Carosati, E.; De Boeck, B.; Ethirajulu, K.; Mackie, C.; Howe, T.; Vianello, R. MetaSite: understanding metabolism in human cytochromes from the perspective of the chemist. *J. Med. Chem.* **2005**, *48*, 6970–6979.
- (19) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849–857.
- (20) Zhou, D.; Afzelius, L.; Grimm, S. W.; Andersson, T. B.; Zauhar, R. J.; Zamora, I. Comparison of methods for the prediction of the metabolic sites for CYP3A4-mediated metabolic reactions. *Drug Metab. Dispos.* **2006**, *34*, 976–983.
- (21) Rydberg, P.; Gloriam, D. E.; Zaretzki, J.; Breneman, C.; Olsen, L. SMARTCyp: A 2D Method for Prediction of Cytochrome P450-Mediated Drug Metabolism. *ACS Med. Chem. Lett.* **2010**, *1*, 96–100.
- (22) Shin, Y. G.; Le, H.; Khojasteh, C.; Hop, C. E. Comparison of Metabolic Soft Spot Predictions of CYP3A4, CYP2C9 and CYP2D6 Substrates Using MetaSite and StarDrop. *Comb. Chem. High Throughput Screening* **2011**, *14*, 811–823.
- (23) Rydberg, P.; Olsen, L. Ligand-Based Site of Metabolism Prediction for Cytochrome P450 2D6. *ACS Med. Chem. Lett.* **2012**, *3*, 69–73.
- (24) Daylight Chemical Information Systems, I., SMARTS - A language for describing molecular patterns, <http://www.daylight.com/dayhtml/doc/theory/smarts.html> (accessed May 31, 2012).
- (25) Guengerich, F. P.; Hanna, I. H.; Martin, M. V.; Gillam, E. M. Role of glutamic acid 216 in cytochrome P450 2D6 substrate binding and catalysis. *Biochemistry* **2003**, *42*, 1245–1253.
- (26) Rowland-Yeo, K.; Rostami-Hodjegan, A.; Tucker, G. T. Abundance of cytochromes P450 in human liver: a meta-analysis. *Br. J. Clin. Pharmacol.* **2004**, *57*, 687–688.
- (27) Williams, J. A.; Hyland, R.; Jones, B. C.; Smith, D. A.; Hurst, S.; Goosen, T. C.; Peterkin, V.; Koup, J. R.; Ball, S. E. Drug-drug interactions for UDP-glucuronosyltransferase substrates: a pharmacokinetic explanation for typically observed low exposure (AUCi/AUC) ratios. *Drug Metab. Dispos.* **2004**, *32*, 1201–1208.
- (28) Ekins, S.; de Groot, M. J.; Jones, J. P. Pharmacophore and three-dimensional quantitative structure activity relationship methods for modeling cytochrome p450 active sites. *Drug Metab. Dispos.* **2001**, *29*, 936–944.
- (29) Sansen, S.; Yano, J. K.; Reynald, R. L.; Schoch, G. A.; Griffin, K. J.; Stout, C. D.; Johnson, E. F. Adaptations for the oxidation of polycyclic aromatic hydrocarbons exhibited by the structure of human P450 1A2. *J. Biol. Chem.* **2007**, *282*, 14348–14355.
- (30) Smith, D. A.; Ackland, M. J.; Jones, B. C. Properties of cytochrome P450 isoenzymes and their substrates part 1: active site characteristics. *Drug Discovery Today* **1997**, *2*, 406–478.
- (31) Smith, D. A.; Ackland, M. J.; Jones, B. C. Properties of cytochrome P450 isoenzymes and their substrates part 2: properties of cytochrome P450 substrates. *Drug Discovery Today* **1997**, *2*, 479–486.
- (32) Rydberg, P.; Vasanthanathan, P.; Oostenbrink, C.; Olsen, L. Fast prediction of cytochrome P450 mediated drug metabolism. *Chem-MedChem* **2009**, *4*, 2070–2079.
- (33) Prasad, J. C.; Goldstone, J. V.; Camacho, C. J.; Vajda, S.; Stegeman, J. J. Ensemble modeling of substrate binding to cytochromes P450: analysis of catalytic differences between CYP1A orthologs. *Biochemistry* **2007**, *46*, 2640–2654.
- (34) Tai, G.; Dickmann, L. J.; Matovic, N.; DeVoss, J. J.; Gillam, E. M.; Rettie, A. E. Re-engineering of CYP2C9 to probe acid-base substrate selectivity. *Drug Metab. Dispos.* **2008**, *36*, 1992–1997.
- (35) Wester, M. R.; Yano, J. K.; Schoch, G. A.; Yang, C.; Griffin, K. J.; Stout, C. D.; Johnson, E. F. The structure of human cytochrome P450 2C9 complexed with flurbiprofen at 2.0-A resolution. *J. Biol. Chem.* **2004**, *279*, 35630–35637.
- (36) Sykes, M. J.; McKinnon, R. A.; Miners, J. O. Prediction of metabolism by cytochrome P450 2C9: alignment and docking studies of a validated database of substrates. *J. Med. Chem.* **2008**, *51*, 780–791.
- (37) Jung, F.; Griffin, K. J.; Song, W.; Richardson, T. H.; Yang, M.; Johnson, E. F. Identification of amino acid substitutions that confer a high affinity for sulfaphenazole binding and a high catalytic efficiency for warfarin metabolism to P450 2C19. *Biochemistry* **1998**, *37*, 16270–16279.
- (38) Lewis, D. F.; Dickins, M.; Weaver, R. J.; Eddershaw, P. J.; Goldfarb, P. S.; Tarbit, M. H. Molecular modelling of human CYP2C subfamily enzymes CYP2C9 and CYP2C19: rationalization of substrate specificity and site-directed mutagenesis experiments in the CYP2C subfamily. *Xenobiotica* **1998**, *28*, 235–268.
- (39) Wada, Y.; Mitsuda, M.; Ishihara, Y.; Watanabe, M.; Iwasaki, M.; Asahi, S. Important amino acid residues that confer CYP2C19 selective activity to CYP2C9. *J. Biochem.* **2008**, *144*, 323–333.
- (40) Ibeau, G. C.; Ghanayem, B. I.; Linko, P.; Li, L.; Pederson, L. G.; Goldstein, J. A. Identification of residues 99, 220, and 221 of human cytochrome P450 2C19 as key determinants of omeprazole activity. *J. Biol. Chem.* **1996**, *271*, 12496–12501.
- (41) Schoch, G. A.; Yano, J. K.; Wester, M. R.; Griffin, K. J.; Stout, C. D.; Johnson, E. F. Structure of human microsomal cytochrome P450

2C8. Evidence for a peripheral fatty acid binding site. *J. Biol. Chem.* **2004**, *279*, 9497–9503.

(42) Scott, E. E.; He, Y. A.; Wester, M. R.; White, M. A.; Chin, C. C.; Halpert, J. R.; Johnson, E. F.; Stout, C. D. An open conformation of mammalian cytochrome P450 2B4 at 1.6-A resolution. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13196–13201.

(43) Williams, P. A.; Cosme, J.; Sridhar, V.; Johnson, E. F.; McRee, D. E. Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity. *Mol. Cell* **2000**, *5*, 121–131.

(44) Locuson, C. W., 2nd; Suzuki, H.; Rettie, A. E.; Jones, J. P. Charge and substituent effects on affinity and metabolism of benzboromarone-based CYP2C19 inhibitors. *J. Med. Chem.* **2004**, *47*, 6768–6776.

(45) Knox, C.; Law, V.; Jewison, T.; Liu, P.; Ly, S.; Frolkis, A.; Pon, A.; Banco, K.; Mak, C.; Neveu, V.; Djoumbou, Y.; Eisner, R.; Guo, A. C.; Wishart, D. S. DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res.* **2011**, *39* (Database issue), D1035–41.